Proskauer Rose LLP | 1001 Pennsylvania Avenue, NW | Washington, DC 20004

Colin R. Kass
202-416-6890
ckass@proskauer.com

April 9, 2025

Hon. William H. Alsup
U.S. Dist. Court, N.D. Cal.

Re: *X Corp. v. Bright Data Ltd., 23-cv-03698 (N.D. Cal.)*[1]

Dear Judge Alsup:

At the last hearing, this Court stressed "the single most important issue" raised by the Amended Complaint "was the extent to which scraping prejudices [X's] server[s]" and whether any servers have been "compromised." March 27, 2025 Hrg. Tr. 63. Despite repeated efforts over ***six months*** to obtain this information from X, X has not identified any server failures that may be at issue or even what information systems X maintains that could bear on this issue. To remedy this, Bright Data requests an (i) order compelling X to produce its responsive server activity logs, reports, and data in accordance with the attached Data Protocol (Ex. 1), and (ii) permission to serve a 30(b)(6) deposition notice relating to the location and existence of X's data systems without counting towards any deposition limits in this case (Ex. 2).

A. The Court Should Order X to Produce its Server Logs, Reports, and Data In Accordance with Bright Data's Server Protocol.

In its Initial Disclosures, X stated it may rely on, among other things, data on "tracking scrapers, blocking, circumvention, and identification;" "X's traffic, rates of inauthentic requests, the volume, the endpoints they target, and the burden;" "what data X collects, how it is collected, what is sold, what is restricted, what is available to non-logged in versus logged in users, what is stored in caches versus cold storage, and any other relevant information on X's data;" and on "server failures, glitches, lag, stolen CPU cycles, and other degraded performance due to inauthentic requests." *See* X's Oct. 31, 2024 Third Suppl. Initial Disclosures.

But X ***has not identified*** the information systems it maintains that contain this data, and still has not provided any indication of what data it will produce. X has also affirmatively stated that it will ***not*** produce the "entirety of [its responsive] server logs." Instead, continuing to stonewall, X says it is just now beginning to "investigat[e] what data it can produce" and that it will, at some point, unilaterally decide what information is "reasonabl[e]" to produce. That approach is not consistent with X's discovery obligations or good faith negotiations.

X's refusal to produce this information is especially unreasonable in light of Bright Data's extensive data productions. Bright Data has completed its production of information from its CRM database, including disclosing over 4,000 customers. Bright Data has also produced sample activity logs, which disclose thousands of specific IP addresses (or "proxies") used to communicate with X's servers, including the type of proxy Bright Data used, and the specific time

[1] Emphasis added, internal citations and quotation marks omitted, and capitalizations conformed without brackets. Bright Data has met and conferred without success on these issues via zoom on February 7th, March 11th, and April 3, 2025; and through written correspondence on February 11th, March 6th, 11th, 13th, 27th and April 1st.

stamps of the communication.[2] Bright Data has also served interrogatory responses quantifying the amount of traffic on its network related to X's platform based on its active data systems.

Now that Bright Data has produced this information, there is no excuse for X to continue to withhold the information on its side of the ledger. Only X has data on the routes Bright Data's requests took within X's server infrastructure, what their impact (if any) was on X's servers, and what roadblocks X tried to put up to block those requests. As such, X needs to produce the following information about its side of the communications with Bright Data:

- ***Platform Engagement Information.*** For each IP address that Bright Data discloses, all information about how the IP address was used to engage with X's platform during the period of engagement reflected in Bright Data's logs *and* at other times. This will include for each session identified: the time stamps associated with the session; all endpoints and URLs visited during the session; any actions taken (including posting or viewing content) during each session; whether any of the endpoints required a log-in to access; whether the user was logged in during the session; the user account, if any, used during the session; whether a CAPTCHA was shown during the session; whether a rate limiter was triggered during the session; any device associated with the session; and whether X issued any error codes, blocks, or other notifications with respect to that IP address, user, or device.

- ***Server Information.*** For each use of an IP address, the identity of the servers accessed, including the name and location of the server; the ownership of the server (*e.g.*, X-owned or cloud service provider); whether there was a server failure or strain at the time of access; and all server performance metrics at the time of such access. X will also need to produce the percentage of server requests attributable to Bright Data or its IP addresses compared to other server requests at the time of any server failure or degradation.

- ***User/Customer Information.*** For each customer and IP address Bright Data discloses, information from X's user account management and customer databases, including information about user accounts associated with Bright Data's customers or IP addresses such as when the accounts were opened or closed; any reports of misuse associated with the accounts (and associate devices or IP addresses); and any business communications with such customers regarding any data sales or licenses.

In an effort to forestall production of this information, X argued it is unnecessary for it to produce its data just because Bright Data produced its data, claiming that equating the two production obligations is a "false equivalence." But one-sided discovery makes no sense. How can X assert that Bright Data committed some violation, yet deprive Bright Data of the information necessary to show that the alleged violation was no violation at all?

Rather than argue relevance, X says that because its claims cover almost five years and over 2,100 days, producing such data would be unduly burdensome for X and unmanageable for Bright Data. As a general matter, Bright Data does not dispute that it *may* be unduly burdensome

[2] During the meet and confer, X argued that Bright Data should produce archived logs for over 2,100 days, but that X need not do the same. As discussed below, *both* sides of a given communication must be produced, with burden addressed via sampling or other unbiased techniques, as Bright Data proposes.

for each party to produce *everything* they have concerning *every* touchpoint between the two companies. That is why Bright Data has proposed a data protocol that relies *in the first instance* on sampling techniques to discover – in a manageable way – what useful information is available and can be reasonably extracted. Under the protocol, the parties would select six baseline sample days, ten alleged server failures allegedly attributed to Bright Data, ten alleged instances of unauthorized scraping or access, and a comparable number of server failures caused by other factors. For each of these days/instances, the parties would produce all data, logs, reports, and documents responsive to the other parties' documents requests. After the parties produce this information, either party may seek additional discovery if needed.

Bright Data's proposed Protocol provides a reasonable process for collecting and matching the relevant information in both parties' information systems. *See* Manual for Complex Litigation, Other Practices to Save Time and Expense § 11.423 (touting the benefits of sampling computer data). The Protocol is also consistent with this Court's Standing Order, which notes that "the goal … is to reduce the overall burden while still locating materials whose relevance and importance justifies the burden." *Standing Order*, ¶ 18.

Nor has X offered any reason why this protocol would not work. To the contrary, X ***agrees*** that a data protocol is an appropriate way to manage burden. In its own discovery responses, X asserts that a "subset" of its "logs and data" "would be sufficient." And, in a March 10th letter, X said it would "extract data … pursuant to [an] agreed upon protocol for the production of data." But because X has refused to negotiate the terms of any data protocol or make any commitments concerning such productions, we need this Court's assistance to impose a Data Protocol as a reasonable starting point for discovery.

B. The Court Should Allow Bright Data to Take a 30(b)(6) Custodian of Records Deposition About X's Data Systems Separate from any Deposition Limits.

Over a year ago, Bright Data sought information concerning "X's tracking capabilities and relevant data infrastructure." *See* RFPs 37-40. For each request, X committed to producing responsive documents. *See id.* And after X amended its complaint, Bright Data made its requests for information about X's data infrastructure even more specific. *See* RFP 63. In response, X acknowledged that Bright Data previously requested data infrastructure information in its First RFPs, and again committed to conducting a targeted search for information responsive to these requests *as written*. However, to date, X has still not produced responsive information, which has hampered Bright Data's discovery efforts. As such, Bright Data believes that a 30(b)(6) deposition is appropriate to obtaining the requisite disclosures. Bright Data, however, is mindful of the limits on depositions and 30(b)(6) topics in the Court's Standing Order, and is concerned that, given the complexity of the issues in this case, additional depositions will be needed. Accordingly, Bright Data requests permission to take this records custodian deposition without reducing the number of merits depositions it may later take.

Sincerely,
/s/ Colin R. Kass
Counsel for Bright Data, Ltd.